

# Predicting the brightness of unrelated self-luminous stimuli

Martijn Withouck,<sup>1,\*</sup> Kevin A.G. Smet,<sup>1</sup> Wouter R. Ryckaert,<sup>1</sup> Geert Deconinck<sup>2</sup> and Peter Hanselaer<sup>1</sup>

<sup>1</sup>KU Leuven, ESAT / Light & Lighting Laboratory, Gent, Belgium

<sup>2</sup>KU Leuven, ESAT / ELECTA, Leuven, Belgium

\*[martijn.withouck@kuleuven.be](mailto:martijn.withouck@kuleuven.be)

**Abstract:** In a magnitude estimation experiment, twenty observers rated the brightness of several unrelated, self-luminous stimuli surrounded by a dark background. The performance of a number of existing vision models, color appearance models and models based on the concept of equivalent luminance in predicting brightness has been investigated. Due to a severe underestimation of the Helmholtz-Kohlrausch effect, none of the models performed acceptable. Increasing the weight of the colorfulness contribution to the brightness attribute in the CAM97u model results in a very good correlation between the model predictions and the visually perceived brightness. Finally the experimental results and the brightness prediction from the modified model CAM97u,m are verified through a matching experiment and a validation magnitude estimation experiment.

©2014 Optical Society of America

**OCIS codes:** (330.1690) Color; (330.5020) Perception psychology; (330.5510) Psychophysics.

---

## References and links

1. M. D. Fairchild, *Color Appearance Models*, Second ed., Wiley-IS&T Series in Imaging Science and Technology (John Wiley & Sons Ltd, 2005).
2. R. W. G. Hunt and M. R. Pointer, *Measuring colour*, Fourth ed., Wiley-IS&T Series in Imaging Science and Technology (John Wiley & Sons Ltd, 2011).
3. R. W. G. Hunt, *Measuring colour*, Third ed. (Fountain Press, 1998), pp. 239–246.
4. S. L. Guth, “ATD01 model for color appearances, color differences and chromatic adaptation,” in *9th Congress of the International Colour Association, Proceedings of SPIE Vol. 4421*, 2002).
5. C. Fu, C. Li, G. Cui, M. R. Luo, R. W. G. Hunt, and M. R. Pointer, “An investigation of colour appearance for unrelated colours under photopic and mesopic vision,” *Color Res. Appl.* **37**(4), 238–254 (2012).
6. CIE, *A colour appearance model for colour management systems: CIECAM02*, (CIE Central Bureau, Austria, 2004).
7. N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, and T. Newman, “The CIECAM02 Color Appearance Model,” in *10th Color Imaging Conference, IS&T and SID*, (Scottsdale, Arizona, 2002).
8. Y. Nayatani, “Simple estimation methods for the Helmholtz–Kohlrausch effect,” *Color Res. Appl.* **22**(6), 385–401 (1997).
9. CIE, *Supplementary System of Photometry*, (CIE Central Bureau, 2011).
10. H. Yaguchi and M. Ikeda, “Subadditivity and superadditivity in heterochromatic brightness matching,” *Vision Res.* **23**(12), 1711–1718 (1983).
11. G. Wyszecki and W. S. Stiles, *Color Science*, Second ed. (Wiley, 1982), pp. 410.
12. CIE, *International Lighting Vocabulary*, (CIE Central Bureau, 2011).
13. M. Withouck, K. A. G. Smet, W. R. Ryckaert, M. R. Pointer, G. Deconinck, J. Koenderink, and P. Hanselaer, “Brightness perception of unrelated self-luminous colors,” *J. Opt. Soc. Am. A* **30**(6), 1248–1255 (2013).
14. S. A. Fotios and C. Cheal, “A comparison of simultaneous and sequential brightness judgements,” *Lighting Res. Tech.* **42**(2), 183–197 (2010).
15. ASTM International, *Standard Test Method for Unipolar Magnitude Estimation of Sensory Attributes*, (ASTM International, 2012).
16. P. A. García, R. Huertas, M. Melgosa, and G. Cui, “Measurement of the relationship between perceived and computed color differences,” *J. Opt. Soc. Am. A* **24**(7), 1823–1829 (2007).
17. B. Koo and Y. Kwak, “Color appearance and color connotation models for unrelated colors,” *Color Res. Appl.*, (2013), doi: 10.1002/col.21857.

18. M. R. Luo, A. A. Clarke, P. A. Rhodes, A. Schappo, S. A. R. Scrivener, and C. J. Tait, "Quantifying colour appearance. Part I. Lutchi colour appearance data," *Color Res. Appl.* **16**(3), 166–180 (1991).
  19. IBM SPSS Statistics for Windows, IBM Corp., Armonk, NY, 2012.
  20. A. Field, *Discovering statistics using SPSS*, Third ed. (SAGE, 2009).
  21. S. Fotios, K. Houser, and C. Cheal, "Counterbalancing needed to avoid bias in side-by-side brightness matching tasks," *J. Illum. Eng. Soc.* **4**, 207–223 (2008).
  22. A. Tversky and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* **185**(4157), 1124–1131 (1974).
  23. G. A. Gescheider, "Psychophysical scaling," *Annu. Rev. Psychol.* **39**(1), 169–200 (1988).
- 

## 1. Introduction

Unrelated colors are colors perceived to belong to areas seen in isolation from any other colors [1, 2]. A self-luminous stimulus surrounded by a dark background, like a traffic or marine signal light viewed during a dark night, is a typical example of an unrelated color. Some vision models have been developed to predict the perception of brightness of these kind of stimuli:

- CAM97u, a color appearance model (abbreviated as CAM hereafter) for unrelated colors developed by Hunt [3], allows for the calculation of the perceptual attributes brightness, hue, colorfulness, and saturation under varying conditions by taking into account some of the physiological processes that occur in the human visual system.
- ATD01, a color vision model developed by Guth [4], is developed to predict the brightness, saturation and hue of unrelated colors by transforming the XYZ tristimulus values of a stimulus into LMS cone responses. These LMS cone responses are gain-controlled and undergo a second transformation to yield an achromatic (A) and two chromatic (T and D) signals which are used to calculate the perceptual attributes.
- CAMFu, a CAM for unrelated colors developed by Fu et al. [5], is highly inspired by CAM97u and CIECAM02 [6, 7] and predicts the brightness, hue, colorfulness and saturation.
- $L_{Eq,Nay}(VCC)$ , a model developed by Nayatani [8], is based on the concept of equivalent luminance which is defined as the photopic luminance of a reference stimulus that matches the test stimulus in terms of brightness [9]. The Variable-Chromatic-Color (VCC) method of Nayatani is used when the luminance of the colored stimulus is changed in order to match the achromatic reference.
- $L_{Eq,Nay}(VAC)$ , a similar model of Nayatani as the one described above [8], in which the luminance of the reference achromatic color is changed in order to match the colored stimulus (Variable-Achromatic-Color or VAC method).
- $L_{Eq,CIE}$ , a model developed by the CIE [9], is also based on the concept of equivalent luminance and can be seen as an international agreed supplementary system of photometry.

The Helmholtz-Kohlrausch effect (abbreviated as the H-K effect hereafter) states that highly saturated colors appear brighter than colors having low saturation, even when they are equal in luminance [8–12]. In a previous study [13] the performance of the six vision models described above was evaluated with respect to their ability to include the H-K effect. However, none of the models were fully able to predict the perceived brightness of unrelated stimuli. Especially notable was the failure of CAM97u as it was specifically designed to deal with unrelated colors.

Note that the models based on the concept of equivalent luminance apply in principle only to related colors. However as they take into account the H-K effect, they have been used in this study.

## 2. Brightness prediction in CAM97u

The input parameters of CAM97u are the XYZ tristimulus values of the stimulus and the conditioning field and the photopic and scotopic luminance of the stimulus, the adapting field and the conditioning field. As there is no conditioning field present in our experimental setup, the equi-energy stimulus  $S_E$  is used as the conditioning field and its luminance is taken to be the same as the one of the adapting field.

According to CAM97u, the brightness is calculated using Eq. (1):

$$Q_{CAM97u} = [(1.1)(A_{CAM97u} + w_M \times M_{CAM97u})]^{0.9}. \quad (1)$$

Where  $Q_{CAM97u}$  is the predicted brightness,  $A_{CAM97u}$  the achromatic signal,  $w_M$  the colorfulness weighting factor equal to 0.01 and  $M_{CAM97u}$  the colorfulness [3]. The inclusion of  $M_{CAM97u}$  in the expression for brightness represents the H-K effect.

In a previous study with unrelated, self-luminous, colored stimuli having a constant luminance of 51 cd/m<sup>2</sup> [13], a strong dependence of the observed brightness with the colorfulness  $M_{CAM97u}$  suggested that an increase of the colorfulness weighting factor  $w_M$  in Eq. (1) might lead to a much better brightness prediction. However the number of stimuli considered was too limited to allow for the determination of an improved weighting factor. In this study, a magnitude estimation experiment including colored stimuli having the same luminance,  $L_{10} = 6.23$  cd/m<sup>2</sup>, as well as achromatic stimuli having a luminance between 7.54 cd/m<sup>2</sup> and 47.60 cd/m<sup>2</sup> is described. The results have been used to modify the brightness prediction of CAM97u by increasing the colorfulness weighting factor. To verify the modified model, a matching experiment and a validation magnitude estimation experiment, using random stimuli within a wide range of chromaticity coordinates and luminance values, have been performed.

## 3. Experimental setup

In this study, a viewing room of 3 m wide by 5 m long by 3.5 m high with black walls, a grey ceiling and a grayish black floor carpet was used (see Fig. 1 (left)). A circular stimulus with a diameter of 37 cm, from which the color and luminance was changed by controlling the intensity of red, green and blue LEDs, is presented to the observers at a distance of 211 cm, providing a field of view (FOV) of approximately 10°. The stimulus is surrounded by a dark background (see Fig. 1 (right)). All colorimetric and photometric quantities were calculated using the CIE 10° observer and determined from spectral measurements using a spectroradiometer (MS260i Oriel instruments spectrograph) and a suitable calibration. More details about this setup can be found in [13].



Fig. 1. (left) Experimental setup. (right) Example of a stimulus under dark viewing conditions.

To improve the brightness prediction of CAM97u given by Eq. (1), in particular taking into account the H-K effect correctly, colored stimuli each having more or less the same value of  $A_{CAM97u}$  are preferred, Eq. (1). Therefore a set of 58 colored stimuli with a FOV of 10° and an equal luminance of 6.23 cd/m<sup>2</sup> (standard deviation 0.11 cd/m<sup>2</sup>) have been

selected. Their CIE 1976  $u'_{10}, v'_{10}$  chromaticity coordinates are illustrated in Fig. 2 (left). In addition to these 58 colored stimuli, a set of 17 achromatic stimuli with a luminance between  $7.54 \text{ cd/m}^2$  and  $47.60 \text{ cd/m}^2$  and a chromaticity close to that of illuminant D65 ( $u'_{10}, v'_{10} = 0.1979, 0.4695$ ; mean  $\Delta E_{u',v'} = 0.005$ ) has been selected (see Fig. 2 (right)). The colorfulness of these achromatic stimuli is approximately the same. These stimuli are necessary to obtain a single CAM97u brightness scale appropriate to both chromatic and achromatic stimuli.

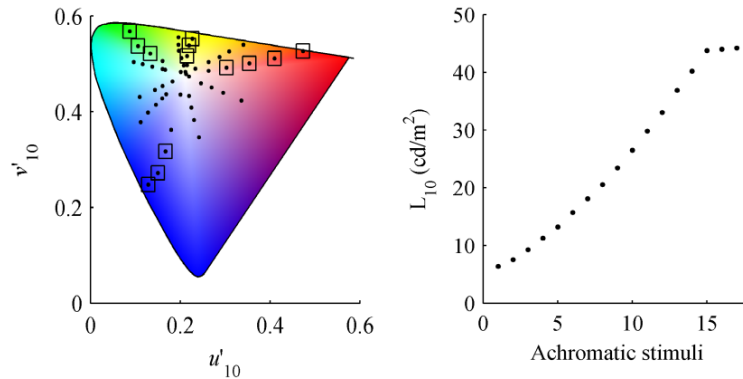


Fig. 2. (left) CIE 1976  $u'_{10}, v'_{10}$  chromaticity coordinates of the 58 colored stimuli. The stimuli highlighted with a squared symbol are also used in a validation experiment. (right) Luminance of the 17 achromatic stimuli calculated using the CIE  $10^\circ$  observer.

#### 4. Visual tests

In a psychophysical experiment, the magnitude estimation method was used in which observers were asked to rate test stimuli in comparison with a reference achromatic stimulus shown in temporal juxtaposition and with a brightness value of 50 attributed to it. The luminance of the reference stimulus was approximately the same as the luminance of the colored stimuli,  $6.38 \text{ cd/m}^2$ , and the chromaticity was close to that of illuminant D65 ( $\Delta E_{u',v'} = 0.006$ ). The experiment started by viewing the reference stimulus. After 5 seconds a test stimulus was shown for 15 seconds. Just after switching back to the reference, again for 5 seconds, the observers were asked to rate the brightness of the test stimulus relative to the reference achromatic stimulus. Before the experiment, the observers adapted to the dark viewing conditions for at least 5 minutes.

The following instructions were given to each observer:

You will see 90 test stimuli. First a reference stimulus will be shown for 5 seconds. Each test stimulus is then presented for 15 seconds. Between each of these 90 test stimuli, the reference stimulus will be shown again for 5 seconds. Give a value to the brightness of the test stimulus immediately after the disappearing of this test stimulus and in comparison with the reference. The reference has a brightness value of 50. A value of zero represents a dark stimulus without any brightness. There is no upper limit to the value of brightness, a value of 100 represents a stimulus appearing double as bright as the reference, a value of 25 is given to a stimulus appearing half as bright, etc.

In this magnitude estimation experiment, 90 test stimuli were presented; 5 stimuli as 'warming up', 75 stimuli as described above and finally 10 stimuli used to calculate observer variability. The stimuli were randomly arranged in two series, each being evaluated by half of the observers to avoid possible biases due to the series sequence [14].

Twenty observers, 10 male and 10 female, with ages ranging between 20 and 31 years (average 25) participated in the psychophysical experiment. All had normal color vision according to the Ishihara 24 plate Test for Color Blindness. Six observers already

participated in previous experiments while the others were naïve with respect to the purpose of the experiment. The naïve observers participated in a training of 45 minutes to become familiar with the magnitude estimation method. They completed a straightforward exercise in which they were asked to rate the length of a line in comparison with a line of length 100, similar to a method described in the ASTM International standard test method for unipolar magnitude estimation of sensory attributes [15]. In addition, a set of training stimuli having the same hue and luminance as being used in the experiments, was also presented, allowing the naïve observers to be aware of the brightness range and to become familiar with the brightness rating technique. The six experienced observers started the magnitude estimation experiment with a set of 25 training stimuli. After a small break the actual experiment started and lasted for approximately 30 minutes.

## 5. Results

### 5.1 Observer variability

The agreement between any two sets of data can be analyzed using the coefficient of variation (CV), Eq. (2) [16]. For a perfect agreement between two sets of data, the CV should be equal to zero. The inter-observer agreement was assessed by calculating the CV values between each individual observer's results and the geometric mean of all the observers [16]:

$$CV = 100 \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(Q_{\text{geom},i} - fQ_{\text{obs},i})^2}{\bar{Q}_{\text{geom}}^2}} \quad (2)$$

$$\text{with } f = \frac{\sum_{i=1}^n Q_{\text{geom},i} Q_{\text{obs},i}}{\sum_{i=1}^n Q_{\text{obs},i}^2}.$$

Where  $Q_{\text{obs},i}$  represents the individual observer brightness of stimulus  $i$ ,  $Q_{\text{geom},i}$  the geometric mean of  $Q_{\text{obs},i}$  for all the observers,  $\bar{Q}_{\text{geom}}$  the arithmetic mean of  $Q_{\text{geom},i}$  over all stimuli,  $n$  the number of evaluated stimuli and  $f$  the scale factor adjusting the  $Q_{\text{geom},i}$  and  $Q_{\text{obs},i}$  values to the same scale.

In addition to the agreement between observers, both short term and long term observer accuracy was also assessed. The short term intra-observer agreement was analyzed by randomly selecting ten stimuli and having each observer rate them a second time at the end of the experiment. This short term agreement was quantified by calculating the CV values between each individual observer's results of the ten stimuli during the test and their results of the same ten stimuli at the end of the test. The observers were not aware that ten stimuli were presented a second time. The long term intra-observer agreement was analyzed by having three observers (2 male, 1 female) repeat the experiment three months later and was quantified by calculating the CV values between each individual observer's results of both experiments. Observers were not told that the experiment was identical.

The results for the inter and intra observer agreement are summarized in Table 1 in terms of CV values. These results show that the mean CV values for inter-observer, short term intra-observer and long term intra-observer agreement are 13%, 11% and 8%, respectively. The CV values are fairly low for all observers, indicating a good agreement.

The CV values are much better than the value of 29% reported by Fu et al. [5] and 40% reported by Koo and Kwak [17] and similar to the value of 11% reported by Withouck et al. [13], when scaling the brightness of unrelated colors in conditions similar to those used in this study. This result is also similar to the inter-observer value of 13% reported by Luo et al. [18] for the lightness of related colors. The CV value for the short term intra-observer agreement of 11% could only be compared to the 15% 'repeatability' obtained by Fu et al. [5], as none of the other studies reported an intra-observer agreement.

**Table 1. Evaluation of Inter-Observer Agreement, Short Term Intra-Observer Agreement and Long Term Intra-Observer Agreement by Calculation of the Coefficient of Variation CV (%).**

	Inter-Observer Agreement		Short Term Intra-Observer Agreement		Long Term Intra-Observer Agreement	
	Male	Female	Male	Female	Male	Female
	12	10	8	9	6	-
	12	11	13	6	-	-
	15	11	19	2	-	-
	13	10	10	13	-	-
	9	12	10	17	-	-
	13	23	13	10	-	-
	9	8	12	7	-	-
	10	12	11	11	7	12
	25	15	11	12	-	-
	23	18	21	10	-	-
Mean	<b>13</b>		<b>11</b>		<b>8</b>	
Median	<b>12</b>		<b>11</b>		<b>7</b>	

### 5.2 Brightness perception

As the same group of observers rated the brightness of each stimulus, a one-way repeated-measures design of ANOVA on all colored stimuli was calculated. The analysis showed that, although all colored stimuli have the same luminance, their brightness perception was significantly different between each other,  $F(1.824, 34.659) = 14.801$ ,  $p < 0.001$ . In Fig. 3, the geometric mean of  $Q_{\text{obs}}$  of all the observers,  $Q_{\text{geom}}$  (“average observer”), are plotted against the *saturation* ( $s_{u'v',10}$ ) of each stimulus, calculated using Eq. (3):

$$s_{u'v',10} = 13 \left[ (u'_{10} - u'_{n,10})^2 + (v'_{10} - v'_{n,10})^2 \right]^{1/2}. \quad (3)$$

Where  $u'_{10}, v'_{10}$  and  $u'_{n,10}, v'_{n,10}$  are the CIE 1976 chromaticity coordinates for the CIE 10° observer of the test stimulus and the reference stimulus respectively.

Figure 3 and Table 2 suggest that for each hue series, all having the same luminance, the perceived brightness increases with saturation, illustrating the Helmholtz-Kohlrausch effect. In fact the slopes of most hue series seem to be coincident, suggesting that the effect of saturation on brightness is equal for all hue series, except for red and blue. Indeed, a customized ANCOVA with  $Q_{\text{geom}}$  as dependent variable, the 11 hues as fixed factors and  $s_{u'v',10}$  as covariate, showed that the effect of the interaction term between  $Q_{\text{geom}}$  and  $s_{u'v',10}$  is significant,  $F(10,1058) = 2.155$ ,  $p < 0.05$ , while the same analysis with only 9 hues (without red and blue) as fixed factors is not significant,  $F(8,842) = 0.485$ ,  $p = 0.867$ . This indicates that the regression slopes are homogeneous for all colors except red and blue. Although some studies reported that the H-K effect is different or even absent for yellow [11], this extensive study suggests that the H-K effect, which is clearly visible, is only different for red and blue.

Remarkably, four of the twenty observers, although obtaining good results in the Farnsworth Munsell 100 Hue Test, rated the red stimuli as being less bright compared to the reference stimulus. Although the other colors were rated brighter and their CV values were normal, these four observers indicated, independently from each other, having trouble with rating the brightness of red stimuli. However by converting the brightness values to z-scores using SPSS [19, 20], it seemed that none of these observer results for the red stimuli are outliers, so they were not removed from the experiment.

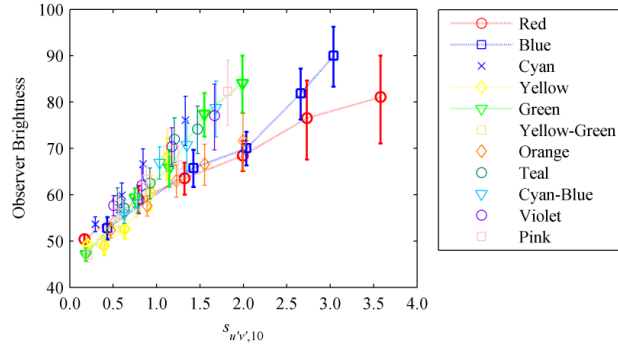


Fig. 3. 'Average observer' brightness ( $Q_{\text{geom}}$ ) with SE error bars, calculated for each individual colored stimulus from all observer answers, plotted against the CIE 1976  $u'_{10}, v'_{10}$  saturation ( $s_{u'v',10}$ ).

**Table 2. Values of 'Average Observer' Brightness ( $Q_{\text{geom}}$ ) and the CIE 1976  $u'_{10}, v'_{10}$  Saturation ( $s_{u'v',10}$ ) Ordered by Hue.**

Color	$Q_{\text{geom}}$	$s_{u'v',10}$	Color	$Q_{\text{geom}}$	$s_{u'v',10}$	Color	$Q_{\text{geom}}$	$s_{u'v',10}$	
Red	81.	3.58	Yellow-Green	69.5	1.1	Violet	77.	1	
	1			3	1		.67		
	76.	2.73		0.9	70.		1		
	6			3	4		.17		
	68.	1.99		0.7	62.		0		
	5			7	0		.83		
	63.	1.32		0.5	57.		0		
	6			6	7		.51		
	59.	0.79		0.3	D65		50.	0	
	0			6	0		.00		
50.	0.17	Orange	71.7	2.0	White	48.	0		
4			0	8		.14			
Blue	90.		3.04			1.5	Ill A	58.	1
	0			5		8	.01		
	81.		2.66	1.2		4000K	51.	0	
	9			3		8	.54		
	70.		2.03	0.8		Achromatic	54.	0	
	0			9			0	.01	
	65.		1.42	0.4			60.	0	
	8			7			2	.01	
	52.	0.43	Teal	74.2	1.4		60.	0	
	8			7	7	.01			
Cyan	76.	1.33			1.2		66.	0	
	1			1	2	.01			
	66.	0.84		0.9		70.	0		
	6		2	8	.01				
	59.	0.60	0.6		71.	0			
	9		2	8	.01				
	53.	0.29	0.5		74.	0			
	6		5	9	.01				
	Yellow	67.	1.12	Cyan-Blue	78.7	1.6		76.	0
		9			8	4	.01		
60.		0.93	1.3			83.	0		
5			5		8	.01			
52.		0.63	1.0			81.	0		
6			3		8	.01			
49.		0.40	0.7			88.	0		
0			8		8	.02			
49.		0.21	0.6			92.	0		
2			3		4	.01			
Green	84.	1.99	Pink	82.3	1.8		88.	0	
	1			2	6	.02			

77.	1.55	66.5	1.2	90.	0
4			5	5	.02
65.	1.15	63.3	0.8	101	0
8			8	.2	.01
59.	0.75	57.9	0.5	90.	0
3			6	2	.01
47.	0.19	47.0	0.2	92.	0
1			1	2	.02

### 5.3 Model performances

The brightness predictions according to the six vision models described before were compared to the ‘average observer’ brightness of the stimuli. In Fig. 4, the ‘averaged observer’ brightness,  $Q_{\text{geom}}$ , has been plotted against the predicted brightness for each of these models. To assess the variability in brightness perception explained by each model, the coefficient of determination ( $R^2$ ) of the linear regression between the observed and predicted brightness has been determined. A  $R^2$  close to 1 suggests a good prediction by the model [20]. Although the relation between the observed and the predicted brightness is expected to be linear, the Spearman correlation coefficient [20] has also been calculated. The Spearman correlation coefficient, having a value between  $-1$  (perfect negative correlation) and  $+1$  (perfect positive correlation), is a rank order metric insensitive to the potential nonlinearity of the relation between the observed and predicted values. In Table 3 the statistical results for each model obtained with all stimuli and with the achromatic stimuli only, are summarized.

Table 3 and Fig. 4, the low values of the Spearman correlation coefficient and the low coefficient of determination for all stimuli are striking. It is clear that none of the described models perform satisfactory, in accordance with the conclusion of previous experiments [13]. However, if only the achromatic stimuli are considered (see Table 3, achromatic stimuli), it can be observed that all models perform very well, indicating that the low overall correlation is due to a severe underestimation of the H-K effect. This is also clearly visible in Fig. 4, showing a different slope for the colored and the achromatic stimuli.

**Table 3. Overview of the Correlation Between the ‘Average Observer’ Brightness Data and the Predictions of the Vision Models.**

Model	$R^2$ (All Stimuli)	Spearman $r$ (All Stimuli)	$R^2$ (Achromatic Stimuli)	Spearman $r$ (Achromatic Stimuli)
$Q_{\text{CAM97u}}$	0.474	0.614	0.950	0.953
$Q_{\text{ATD01}}$	0.565	0.625	0.952	0.953
$Q_{\text{CAMFu}}$	0.467	0.568	0.947	0.953
$L_{\text{Eq,Nay}}$ (VAC)	0.525	0.873	0.908	0.953
$L_{\text{Eq,Nay}}$ (VCC)	0.594	0.780	0.908	0.953
$L_{\text{Eq,CIE}}$	0.506	0.861	0.910	0.953



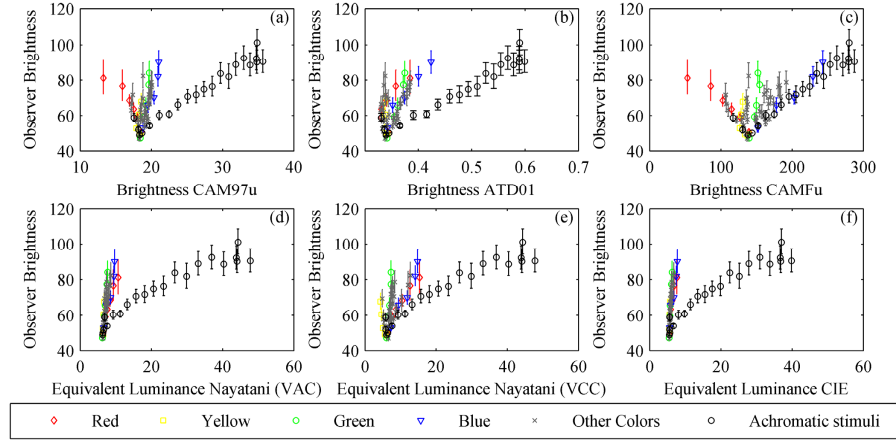


Fig. 4. ‘Average observer’ brightness ( $Q_{\text{geom}}$ ) with SE error bars plotted against the brightness predictions of CAM97u (a), ATD01 (b) and CAMFu (c) and the predictions of the equivalent luminance of Nayatani (VAC (d) and VCC (e)) and CIE (f).

#### 5.4 Modified CAM97u

As suggested before, the brightness prediction of CAM97u, Eq. (1), could be improved by increasing the colorfulness weighting factor  $w_M$ , taking the H-K effect better into account. To determine an optimized colorfulness weighting factor, the ‘average observer’ brightness  $Q_{\text{geom}}$  was first rescaled to the original CAM97u ( $w_M = 0.01$ ) brightness scale but only using the data of the achromatic stimuli for which the CAM97u model works well. This ‘rescaled observer’ brightness,  $Q_r$ , was obtained by multiplying  $Q_{\text{geom}}$  with the slope of the linear regression between  $Q_{\text{geom}}$  and  $Q_{\text{CAM97u}}$  for all 17 achromatic stimuli.

By minimizing the mean of the squared residual errors between  $Q_r$  and the brightness values calculated according to Eq. (1), the value of the colorfulness weighting factor  $w_M$  was optimized from its original value of 0.01 to 0.268. Similar to Eq. (1), the modified brightness model,  $Q_{\text{CAM97u,m}}$ , is given by

$$Q_{\text{CAM97u,m}} = [(1.1)(A_{\text{CAM97u}} + 0.268 \times M_{\text{CAM97u}})]^{0.9}. \quad (4)$$

When plotting the ‘average observer’ brightness against  $Q_{\text{CAM97u,m}}$  for all stimuli, it is clear that the new model outperforms the former models (see Fig. 5). This is confirmed by the high Spearman correlation coefficient (0.961) and the high coefficient of determination (0.914).

A color appearance model can be considered successful when the error of the model’s prediction is smaller than the observer accuracy in terms of inter-observer agreement [5]. The CV coefficients between the observed brightness rating and the predicted brightness results from the color vision models, including the modified model prediction  $Q_{\text{CAM97u,m}}$ , have been calculated using Eq. (2). The CV values range between 14 and 59 for the six original models and are higher than the inter-observer agreement of 13. However the modified brightness model results in a CV of only 6, which proves again his excellent performance.

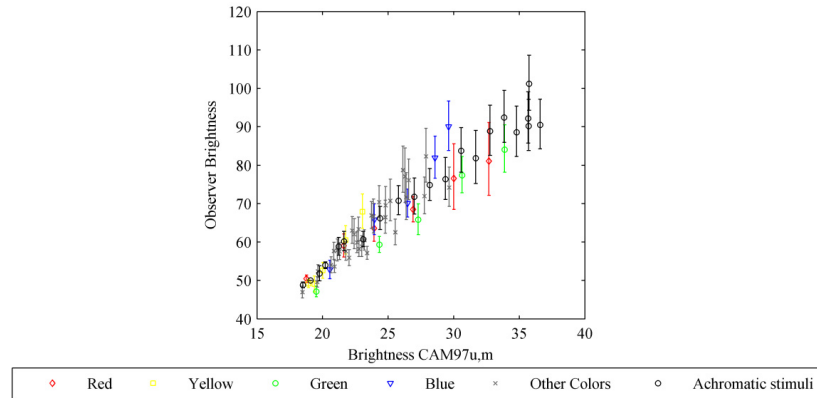


Fig. 5. ‘Average observer’ brightness ( $Q_{\text{geom}}$ ) with SE error bars plotted against the modified brightness predictions of CAM97u,m.

## 6. Validation

### 6.1 Matching experiment

The performance of the modified CAM97u,m model is verified by a successive matching experiment performed by the same observers and using the same experimental setup. The matching experiment started immediately after the magnitude estimation experiment except for a break of ten minutes. In the experiment, which lasted for about 25 minutes, observers adjusted the intensity of the achromatic reference stimulus until it matched that of the colored stimulus in terms of brightness. For the colored stimuli, only the most saturated red (4) and the most saturated blue (3), yellow (3) and green (3) stimuli have been used (see Fig. 2 (left)). The initial luminance of the reference stimulus, shown in temporal juxtaposition with the colored stimuli, was randomly high or low in order to avoid an initial (il)luminance bias [21, 22]. Observers were able to switch back and forth between the reference and the colored stimulus as much as they wanted to until a satisfactory match was found, but as in the magnitude estimation experiment the colored stimulus was always shown for 15 seconds. Two groups of 10 observers viewed the same sequence of colored stimuli but with an opposite initial reference luminance.

The  $10^\circ$  luminance of the reference was measured after each match and an “average matched reference luminance” was obtained for each colored stimulus by taking the mean of all observer matches. A high initial luminance of the reference mostly resulted in a higher matched reference luminance compared with a low initial luminance (see Table 4). This effect is responsible for a luminance difference of 22% between the two experimental conditions. However, the experiment was set up with both conditions having an equal number of matches. By averaging the results, this type of bias will be neutralized [23].

**Table 4. Values of the Matched Reference Luminance of the 13 Colored Stimuli for Both the High and Low Initial Luminance and the Difference Between Them, Ordered by Hue.**

Color	Matched Luminance ( $\text{cd/m}^2$ )		Difference ( $\text{cd/m}^2$ )
	High initial luminance (H)	Low initial luminance (L)	
Red	32.3	29.1	3.1
	23.2	20.3	2.9
	15.9	16.1	-0.3
	15.1	10.4	4.7
Blue	32.1	27.1	5.0
	22.2	25.8	-3.6

	23.2	15.7	7.4
Yellow	15.5	7.9	7.6
	10.7	9.7	1.0
	10.4	6.4	3.9
Green	23.3	18.3	5.1
	20.8	15.7	5.0
	20.1	10.2	9.9
<b>Mean</b>	<b>18.4</b>		<b>4.0</b>

A plot of the averaged matched reference luminance versus the saturation  $s_{uv,10}$  of the stimuli that were to be matched, is given in Fig. 6 (left), clearly illustrating the H-K effect. In Fig. 6 (right) a plot of the modified CAM97u,m brightness of the matched reference against the corresponding stimuli is given. The figure indicates that the modified CAM97u,m brightness is capable to predict the outcome of the matching experiment. The CV value between both brightness predictions was calculated using Eq. (2), using a scale factor  $f$  equal to one because both data sets have the same scale. The low value of 7% confirms the excellent performance of CAM97u,m.

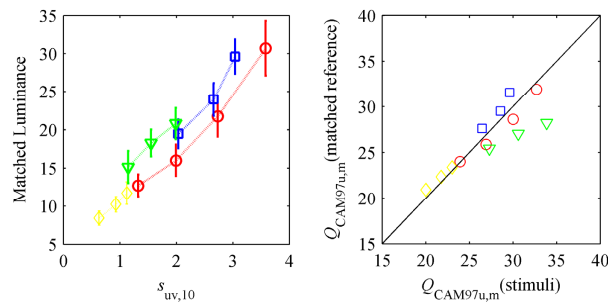


Fig. 6. (left) Matched reference luminance with SE error bars plotted against the CIE 1976  $u'_{10}, v'_{10}$  saturation ( $s_{uv,10}$ ). (right) The modified CAM97u,m brightness prediction of the matched reference plotted against the prediction of the corresponding stimuli.

## 6.2 Magnitude estimation experiment

A decisive magnitude estimation experiment to validate the modified CAM97u,m brightness prediction was set up with 107 stimuli: 15 achromatic stimuli, 40 colored stimuli and 52 'random' stimuli. The luminance of the achromatic stimuli ranged from 5.94 cd/m<sup>2</sup> to 297.47 cd/m<sup>2</sup> (see Fig. 7 (left)) with a chromaticity close to that of illuminant D65 (mean  $\Delta E_{u'v'} = 0.002$ ). The 40 colored stimuli consisted of the four primary hues with both a low and a high saturation, at five luminance levels (see Fig. 7 (left)). The 52 'random' stimuli had a luminance ranged randomly within 6.48 and 57.60 cd/m<sup>2</sup> (see Fig. 7 (middle)) and covered the whole chromaticity gamut of the experimental setup (see Fig. 7 (right)).

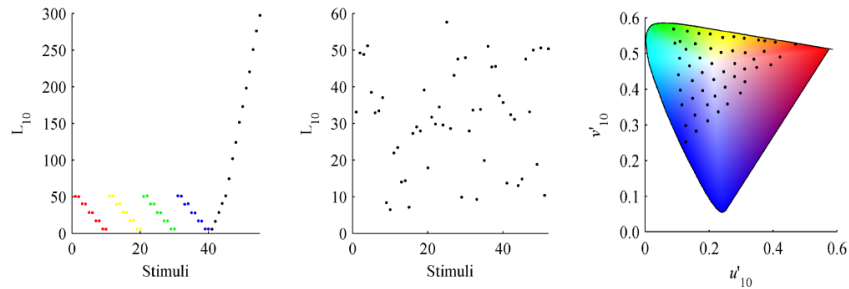


Fig. 7. (left) Luminance values of the 40 colored and the 15 achromatic stimuli. (middle) Luminance values of the 52 'random' stimuli. (right) CIE 1976  $u'_{10}, v'_{10}$  chromaticity coordinates of the 52 'random' stimuli.

The experimental method used in this validation magnitude estimation experiment was identical to the one used in the magnitude estimation experiment described above, except for the reference stimulus to which an intermediate luminance of 43.10 cd/m<sup>2</sup> was attributed. Twenty observers participated in this experiment. All except two had also participated in the first magnitude estimation and matching experiment. The mean CV values for inter-observer, short term intra-observer and long term intra-observer agreement of this validation magnitude estimation experiment are 18%, 12% and 15%, respectively, and are similar to the values mentioned before.

The geometric mean was used to obtain the observer brightness  $Q_{\text{geom}}$  (“average observer”) and was plotted versus  $Q_{\text{CAM97u,m}}$  in Fig. 8. From the coefficient of determination of 0.807, the Spearman correlation coefficient of 0.899, and the CV value of 11%, it is clear that the modified CAM97u,m model gives an excellent prediction for brightness, given by Eq. (4), of colored and achromatic unrelated, self-luminous stimuli covering a wide color gamut and range of luminance levels.

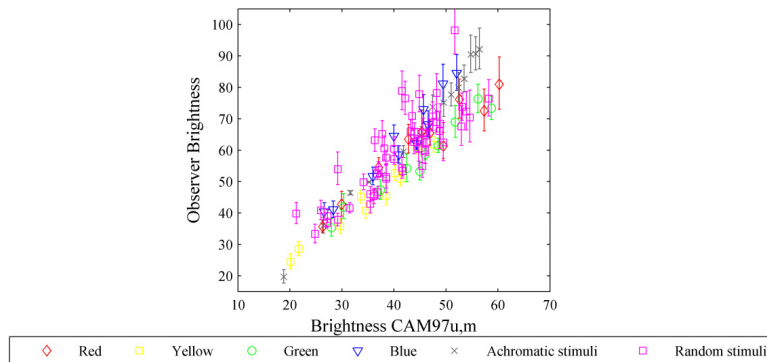


Fig. 8. ‘Average observer’ brightness ( $Q_{\text{geom}}$ ) with SE error bars plotted against the brightness prediction of CAM97u,m.

## 7. Conclusions

The brightness perception of a set of unrelated, self-luminous colored stimuli having a constant luminance of 6.23 cd/m<sup>2</sup>, and of a set of achromatic stimuli, having a luminance ranging from 7.54 cd/m<sup>2</sup> to 47.60 cd/m<sup>2</sup>, was investigated in a magnitude estimation experiment with twenty observers. It was found that the Helmholtz-Kohlrausch effect contributed significantly to the observed brightness. The brightness prediction of six existing vision models was investigated but none of the models performed satisfactory. Adapting the CAM97u model by increasing the colorfulness contribution in the brightness attribute resulted in modified model, called CAM97u,m, which allows for a substantially better brightness prediction.

The performance of the new model was confirmed by both a matching experiment and an extensive validation magnitude estimation experiment using a random sequence of stimuli within a wide chromaticity range, including achromatic ones, and within a wide range of luminance. The modified model CAM97u,m clearly outperformed the other existing vision models and was found to give a reliable brightness prediction for unrelated, self-luminous stimuli.

## Acknowledgements

The authors would like to thank the Research Council of the KU Leuven for supporting this research project (STIM - OT/11/056 and OT/13/069). Author K.S. would also like to thank the Research Foundation Flanders for the support through a postdoctoral fellowship.